IR PROJECT END-SEMESTER EVALUATION

Automatic Medical Image Annotation and Content-based Image Retrieval using Relevance Feedback

By:
Manali A. Shah[15IT218]
Ria Kulshrestha[15IT231]
Shivani Shrivastava [15IT243]
Sneha S. Patil [15IT245]

INTRODUCTION

IMAGE ANNOTATION AND CONTENT BASED IMAGE RETRIEVAL

➔ Image annotation is the task of attaching meta-data in form of caption or keywords to any image.

- Manual Image Annotation: Manual image annotation is the process of manually defining regions in an image and creating a textual description of those regions.
- Automatic Image Annotation: Automatic image annotation (also known as automatic image tagging or linguistic indexing) is the process by which a computer system automatically assigns metadata in the form of captioning or keywords to a digital image.

- Content based Image Retrieval is the content(features) based search for similar images in a large database.
  - Concept-based image indexing, also variably named as "description-based" or "text-based" image indexing/retrieval, refers to retrieval from text-based indexing of images that may employ keywords, subject headings, captions, or natural language text. Indexing is a technique used in CBIR.

LITERATURE REVIEW OUTCOME:

- Despite keyword-based image retrieval providing easier query interface and more accurate retrieval results than CBIR, query results of annotated keywords still is far from user's satisfaction because image annotation is performed based on image visual features.
- MacArthur SD, Brodley CE, Shyu CR: Relevance feedback decision trees in content-based image retrieval. Proceedings of the IEEE Workshop on Content-Based Access of Image and Video Libraries: 68–73, 2000.
    - CBIR uses unweighted K-nearest neighbor retrieval to retrieve the images when no relevance feedback is present. The retrieved images are marked relevant or irrelevant as per the user relevance. This feedback is relayed back to the system and then the algorithm is presented with the K labeled images.

- Ko BC, Kim SH, Nam JY: X-ray image classification using random forests with local wavelet-based CS-local binary patterns. JDigit Imaging, 2011
  - X-ray image classification using random forests with local wavelet-based CS-local binary patterns. It classifies X-ray images using random forests with local wavelet-based local binary pattern (LBP) to improve image classification performance and reduce training and testing time.
- Maria Tzelepi, Anastasios Tefas: Relevance Feedback in Deep Convolutional Neural Networks for Content Based Image Retrieval, 2016
  - Provides relevance Feedback in deep convolutional neural networks. The weights of the CNN are tuned as per the user relevance and the resulting weights are used for retraining. The layers of the CNN which are used for feature extraction are modified so that in the iterations that follow, the new weights are used to change the feature representations so that the relevant images come closer to being similar to the query representation.

PROBLEM STATEMENT

Content-based medical image retrieval and relevance feedback method for image retrieval for enhancing image retrieval performance.

OBJECTIVES

- ➔ Performing image classification based on content of the images.
- ➔ Building an image retrieval system.
- ➔ Provide mechanism for relevance feedback.
- ➔ Improving classification performance based on the relevance feedback.

PROPOSED METHOD

DEEP LEARNING SOLUTIONS FOR MEDICAL IMAGING

- The project is developed using TensorFlow, which is a deep learning software library developed by Google. It is used for high performance numerical computation. The main advantage of using TensorFlow is its flexible architecture which allows easy deployment of computation across various platforms.
- Firstly we re-organized the data so that each folder contains images of the class identified by its by folder name.Then we create a file named label.txt which contains the list of all such folders.
- After that we read the list of folders from label.txt and extracts features like height, width, color space, channels, format, filename, class label along with the image itself and creates a file of the format '.tfrecord', which is simple record-oriented binary format commonly used in Tensorflow application. It also splits the dataset into two sets i.e training and testing. The data is shuffled so that there is a random distribution amongst the two sets.

- The CNN is trained for 5 epochs. It is then used to classify the testing images. The results of the classification are then stored in a CSV file. The CSV file contains the predicted label and the image name.
- For the retrieval system we use the CSV file as reference. When a user enters a particular class number then a list of all the images with the same class in the CSV file is retrieved.
- For feedback we ask the user to enter the image name and the relevant class number. After receiving the feedback, we retrain the model by freezing the convolutional layers and only updating the weights of the fully connected layers.

CNN- what is convolution, activation function, pooling and convolution

- ➔ Convolutional Neural Networks are very similar to ordinary Neural Networks.
- ➔ The convolution and dot product are both linear operators and thus inner products can be written as convolutions and vice versa. By rewriting the fully connected layers as convolutions, the CNN can take input images larger than it was trained on and produce a likelihood map, rather than an output for a single pixel.
- ➔ The resulting 'fully convolutional network' (fCNN) can then be applied to an entire input image or volume in an efficient fashion.

Convolution layer

- ➔ The Conv layer is the core building block of a Convolutional Network that does most of the computational heavy lifting.Three hyperparameters control the size of the output volume: the depth, stride and zero-padding.

Pooling Layer

- ➔ Its function is to progressively reduce the spatial size of the representation to reduce the amount of parameters and computation in the network, and hence to also control overfitting.
- ➔ The Pooling Layer operates independently on every depth slice of the input and resizes it spatially, using the MAX operation. In addition to max pooling, the pooling units can also perform other functions, such as average pooling.

Fully-connected layer

- ➔ Neurons in a fully connected layer have full connections to all activations in the previous layer, as seen in regular Neural Networks.

ARCHITECTURE (Proposed architecture)

- ➔ Number of convolution layers: 5
- ➔ Number of fully connected layers: 4
- ➔ Loss function: Cross-Entropy loss
- ➔ Optimizer: Adam optimizer
- ➔ Learning Rate: 1e-4

RESULTS AND ANALYSIS

- The proposed model was not able to achieve statistically satisfactory results even though the training dataset has more than 10,000 images. An explanation for the same can be found in the small number of images per class which hinders the leaning of the model.
- Number of epochs : 3
- Number of classes : 58
- Batch Size : 120 Accuracy of the model : 40\%

- The metrics for evaluation include :
- k-precision
- k-recall

Since there is a unique occurence of query in the test set, the numerator of k-recall and k-precision will always be 1 if it has correctly been identified by the model else it will be 0.Also, If the image does not exist in the dataset, the value will be infinity.

- For k=5,10, 20 and 40
- Query Entered : 740028.png
- k-precision for k=5 : 0.2
- k-recall for k=5 : 0.33
- k-precision for k=10 : 0.1
- k-recall for k=10 : 0.2
- k-precision for k=20 : 0.05
- k-recall for k=20 : 0.09
- k-precision for k=40 : 0.025
- k-recall for k=40 : 0.04

- Due to the low accuracy of the model, the performance metrics k-precision and k-recall are not giving satisfactory values.Also, since there is an unique occurence of the image in the dataset the numerator for the information retrieval metrics is always going to be 1 or 0.The denominator will increase as we increase the value of k. The accuracy of the CNN model can be improved by training it for more number of epochs. Since there is an unbalanced distribution of data among the classes, we have chosen the top 58 classes and have eliminated the classes with lower number of images to train the dataset.

```
minorprojectgp4@Research: ~/cnn/seperated
loss:
204.4819
372
label_out:
[54 40 21  9 58 38 53  6 26 53 51 48 37 53 32  6 23 49 10 53 43 18 25 39
 11 56 46 33 16 46 18 46 53 56  8 13 53 54  9  3 53 56 54 51  8  6 53 46
 51  2]
infer_out:
loss:
203.4819
373
label_out:
[53  6 53 53  6 53 54  1 51 54 51 46 51 51 53 54 28 53 23 33 51 29 24 53
 43 53 51 54 55 49 29 53 53 33 43 54 53 51 53 53 11 53 53  1 57 56 46 22
 54  8]
infer_out:
loss:
204.4819
374
label_out:
[58  8 58 51 54 16 53 49 54  4  9 46 23 51 40 32 51 54 54 51 11 11 54 25
 32 48 53 53 10 20 53 39 11 36 54 53 54 11 38 19  8 53 29 17  9 32 54  8
 25 37]
infer_out:
loss:
204.4819
375
label_out:
[19  8 51 13 22 51 11 58  9 54  8 49 13 47 15 50 46 53 11  2  2 54 54 24
  9 53 24 53 46 53  1 14 33 29 16  5  3 51 54 11 53 22  6 45 11  5 34 55
  3  6]
infer_out:
loss:
203.4819
376
label_out:
[ 6  9 11 54 51 53 10 44 54  2  6 44 53  8 54 24 14 53 54 35 52 54 33  5
  8  9 23 53 49 30 17 57  6 54 32 38  8 54  6 22  9 29 51 23 34 53 53 15
 52 19]
infer_out:
loss:
203.4819
377
label_out:
[34 11  9 13 29  8  8 54 11  6 53 53 53 25 25 33 45 37 39 53 21  9 25 44
 54 57 53 25  9 54  5 53 58 55 38 46 27 19 53 34 53 53 28 54 51 53 18 26
 51 53]
infer_out:
loss:
204.4819
378
label_out:
[24 24 47  6  8 51 53 54 30 51 47 54 25 54 53 53 33 52 39  1  2 53 10 53
 31 51 54 54 53 51 23 53 47 37 51 24 53 51  2 54 54 22 36 41 33 51 53 54
 51 53]
infer_out:
loss:
204.4819
379
label_out:
[53 53 49 58 51 45 25 34  8 54 51 51 40 54 55 45 39 32 44 54 53 53 36 53
 54 53 51  8 54 31 30 53 11  9 53 57 54 34  8 51 53  8 51 54 35 32  8 49
```

Loss per 50 images

CHALLENGES:

- ➔ Lack of available annotated images.
- ➔ Relevance feedback in CNN.

FUTURE WORK

- Training on the current dataset for 1 epochs for better results.
- Training on deeper architectures (not done as the dataset available was small)
- Training on the datasets containing different types of medical images like MRIs, CT scans, X-Rays etc.